Elimination of the Guessing Factor from Multiple Choice Testing

by Richard G. Harrison

ABSTRACT

Introduced is a computer testing method which compensates for random guessing. The approach, applicable only in uniform instruction situations, improves the reliability of measurement markedly. A typical large-group lecture/large-group test satisfying the uniformity restraint indicates that results are reliable to within one-tenth of a point.

INTRODUCTION

One of the most challenging problems facing a teacher, whether at the grade school or the graduate school level, is the accurate measurement of students understanding of course material. The traditional methods of oral and written examination have been replaced in many cases by computer-scored testing. One of the greatest drawbacks of this new technique is the reward given to guesses. As any teacher knows, however, there are actually two types of guesses: educated guesses and lucky guesses. Since our job is education and our goal is measurement, it is desirable to measure the educated-ness of each student's answer to each question.

Consider an examination group of 5 students, the first student being superior to the second, the second superior to the third, and so on. Consider also an examination consisting of 5 questions, the first question being easier than the second, the second easier than the third, and so on. One might expect the performance on such the examination to be as follows:

Question		1	2	3	4	5	O－correct X－incorrect
Student	Pts	5	4	3	2	1	
A	5	O	O	O	O	O	
B	4	O	O	O	O	X	
C	3	O	O	O	X	X	
D	2	O	O	X	X	X	
E	1	O	X	X	X	X	

Consider, however, that among questions having 4 alternatives, a student should receive credit for 1/4th of the questions simply by randomly selecting alternatives with no consideration of the individual alternatives whatsoever. One would expect that student E would "guess" correctly on 1 of 4 questions for which the correct answer was not known.

It is not always to possible to detect a guessed answer from a known one: e. g., student E has a 25% chance of guessing question 2, which would produce a result pattern of O-O-X-X-X which is identical to that of student D. Yet with a result pattern of O-X-X-X-O one would feel justified in saying that question 5 had been guessed.

Instructors, as with people in general, tend to view things as being either white or black, good or bad, known or guessed: one benefit of computers is the ability to view things in a fractional manner. This benefit can be used to minimize the effect of this disadvantage of computer testing.

APPROXIMATION

Let N_j be the number of students correctly answering the j^{th} question. Without loss of generality, let $N_j > N_{j+1}$. Let $R_{i,j}$ be the i^{th} student's result for the j^{th} question, where the value 1 represents a correct response and the value of 0 represents an incorrect response. Let $P_{i,j}$ be an approximation of the probability of the i^{th} student knowing the correct

answer to the j^{th} question. The value of $p_{i,j}$ can be approximated as

$$P_{i,j}=\frac{\sum_{j'<j}(N_{j'}\times R_{i,j'})+N_j}{\sum_{j'<j}(N_{j'})+N_j}\times R_{i,j}$$

Note that this approximation is directly parallel to the approximation of the probability of a question being known by students, with the score of the i^{th} student being greater than that of the $(i+1)^{st}$ student.

The trivial case ($j=1$) provides the value of 1 — any student, good or poor, has the greatest chance of knowing the correct answer for the very easiest of questions (a trivial question). Later questions provide values which vary according to the students success on previous (easier) questions. In a very real sense, the value measures the cohesion exhibited among such questions.

The variation of this approximation can be seen in Figures IA and IB, where 2 of 19 students, both of whom have correctly answered 8 of 20 questions on a test, are charted according to result cohesion values. Ordering the questions according to correct-answer counts reveals an interesting pattern: Student A has been correct on the easier questions, whereas Student B has been sporadically correct throughout the test, regardless of question difficulty.

Figure IA

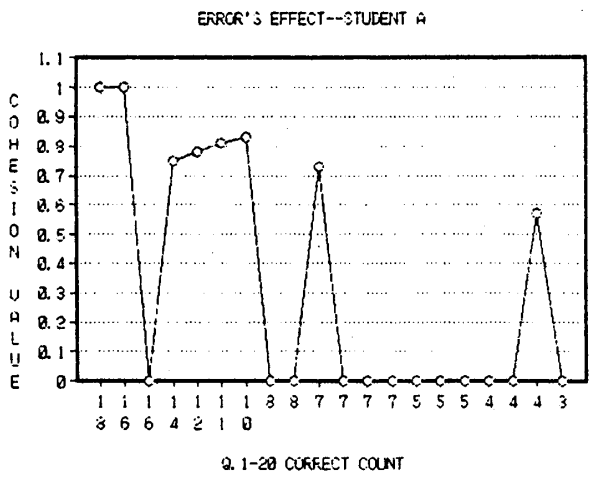


Figure IB

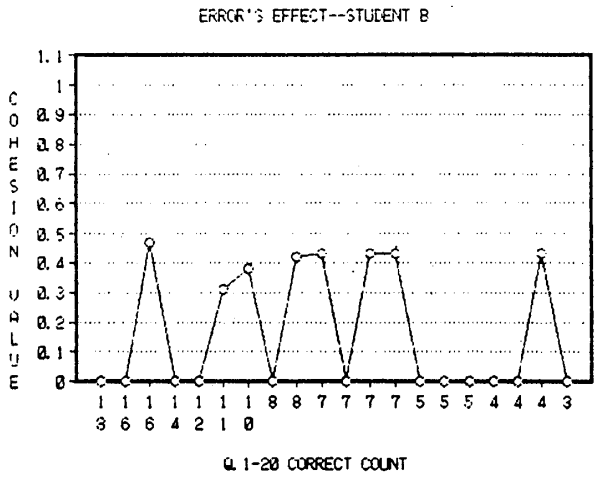


Figure IA shows the cohesion idea. Student A correctly answered the easiest question (Student B was the only student to incorrectly answer this question) for a cohesion value of 18/18. The second easiest question was also answered correctly for a cohesion value of

1=(18+16)/(18+16). The third question, which had been answered 16 times as well, interrupted the cohesion among the questions, so that the fourth question received a cohesion value of 0.76=(18+16+0+14)/(18+16+16+14). The following three questions increased cohesion, with the value $(\sum_{j'\leq j} N_{j'} - N_3)/(\sum_{j'\leq j} N_j)$, which is asymptotic to 0.91.

This same asymptotic-like effect is seen in Figure IB, where sporadically correct answers produced a curve approaching 0.45.

Although it is straightforward how cohesion values can be found, it must be kept in mind that the cohesion sum is primarily a measurement of the degree of accuracy of the score to the knowledge anticipated. The percentage of score which is knowledge might be termed a cohesion quotient, which would be the ratio of the cohesion sum to the previously assigned student score. For a typical test of 19 students and 20 questions, question cohesion and student cohesion calculation lists are shown in Tables IA and IB: the "+" mark indicates a cohesion value of 1, cohesion values of 0.99 to 0.01 are indicated by a truncated

Table IA

QUESTION COHESION

NAME	POINT	CQ	Q1	Q3	Q2	Q4	Q5	Q7	Q9	Q8	Q6	Q9	Q0	Q5	Q7	Q4	Q8	Q6	Q0	Q2	Q1	Q1	Q3
	POINT		18	16	16	14	12	10	9	8	8	7	7	7	7	5	5	5	4	4	4	3	
オヲ	13		+	+	+	+	+		+	+	+		+	+				+		+	+		
イウ	12		+	+	+	+		4	+	+	+				4		4			+			4
ヒ	12		+	+	+	+	+				+	4	+				4	+		+			4
I	10		+	+	+		7	4	7		+		7	4	4								
オオ	10		+	+	+	+	7	4	7		+				4	2							
アオ	10		+	+	+		7	4	7	7	+	4							2				
ゴ"ウ	10		+	+	+	+	7	4	7			4				2			2				
オ	9		+	+	+	7								3		3	3	3	3				
カ"ワ	9		+	+	+					8		8		5	3		3		3				
キ	9		+	+	+	7		7			5			5		3		3					
ウラ	8		+	+	+	7	6	6			4			5									
タ	8				+				6	7		7	3		4	3							2
オノ	8		+	+		7	6	6	7				3									1	
カワ	8		+	+	+	7					4		3		4	3							
アイ	7		9	9	9	7	6									3			3				
イ	7		9	9	9	7	6							4	4								
ウI	7		9			7		6									3	3			3	1	
オカ	7		9			7	6	6			4		4									1	
ゴ"	7		9	9	9		6		6	4						3							
	COHESION		9	9	9	8	7	5	7	6	9	4	6	5	4	2	4	6	2	8	3	4	
			8	8	8	3	2	5	9	3	5	1	8	0	0	7	1	2	7	2	9	1	

tenths-of-a-point digit, and incorrect answers are indicated by blank spaces. Cohesion values have been processed with two decimal digits of accuracy (truncated), while cohesion quotients and scores have been expressed as whole numbers (rounded).

Table IB

STUDENT COHESION

NAME	POINT	CQ	Q1	Q2	Q3	Q4	Q5	Q7	Q9	Q8	Q6	Q9	Q0	Q5	Q7	Q4	Q8	Q6	Q0	Q2	Q1	Q1	Q3
	POINT		1	1	1	1	1	1	1														
			8	6	6	4	2	1	0	8	8	7	7	7	7	5	5	5	4	4	4	3
	COHESION		9	9	9	8	7	5	7	6	9	4	6	5	4	2	4	6	2	8	3	4
			8	8	8	3	2	5	9	3	5	1	8	0	0	7	1	2	7	2	9	1
ｵｦ	13	91	+	+	+	+	+		8	8	8		9	9				8		7	7		
ｲﾃﾞ	12	88	+	+	+	+		8	8	8	8				9		7			7		7	
ﾋ	12	84	+	+	+	+	+				8	7	7				7	7		7		6	
I	10	88	+	+	+		8	8	8		8		8	8	8								
ｵｵ	10	97	+	+	+	+	+	+	+		+				9	8							
ｱｵ	10	88	+	+	+		8	8	8	8	8	8							6				
ｺﾞｳ	10	93	+	+	+	+	+	+	+			8					7			7			
ｵ	9	75	+	+	+	+								5			5	5	5	5			
ｶﾞﾜ	9	74	+	+	+				6		6		6	6				5		5			
ｷ	9	86	+	+	+	+		8		7			7			6		6					
ｳﾗ	8	97	+	+	+	+	+	+		9			8										
ｿ	8	41			4			3	3		4	4		4	4								4
ｵﾉ	8	81	+	+		7	7	8	8			7										5	
ｶﾜ	8	83	+	+	+	+				6		6		6	6								
ｱｲ	7	90	+	+	+	+	+								6			6					
ｲ	7	91	+	+	+	+	+						6	6									
ｳI	7	48	+			5		4									3	3			3	3	
ｵｶ	7	62	+			5	5	6		6		5										4	
ｺﾞ	7	86	+	+	+		8		7	7					7								

CALCULATION

The cohesion sum is simply the sum of the approximations of knowledge probability for student- or question-responses. In order to confirm the applicability of this approach, cohesion sums were re-assigned as question and student "scores", and the cohesion measurement process repeated, and then the entire process was repeated again. The entire measurement process was applied to a forty-question test, with student sample sizes of 19, 39, and 99. Results for the 19 student group are shown in Table II, and indicate that the approach does agree with itself: cohesion quotients (score reliability) jumped from an 80% average to a 99% average, and then remained at that high level. Among students in the original test sample of 19, cohesion quotients showed a standard deviation of 10%, which

dropped to 4-5%. and further improved to 2.5-3%. Given the examiner's natural tendency toward whole (integer) values, whereas cohesion values are by definition fractions (real numbers), such deviation between values is justifiable.

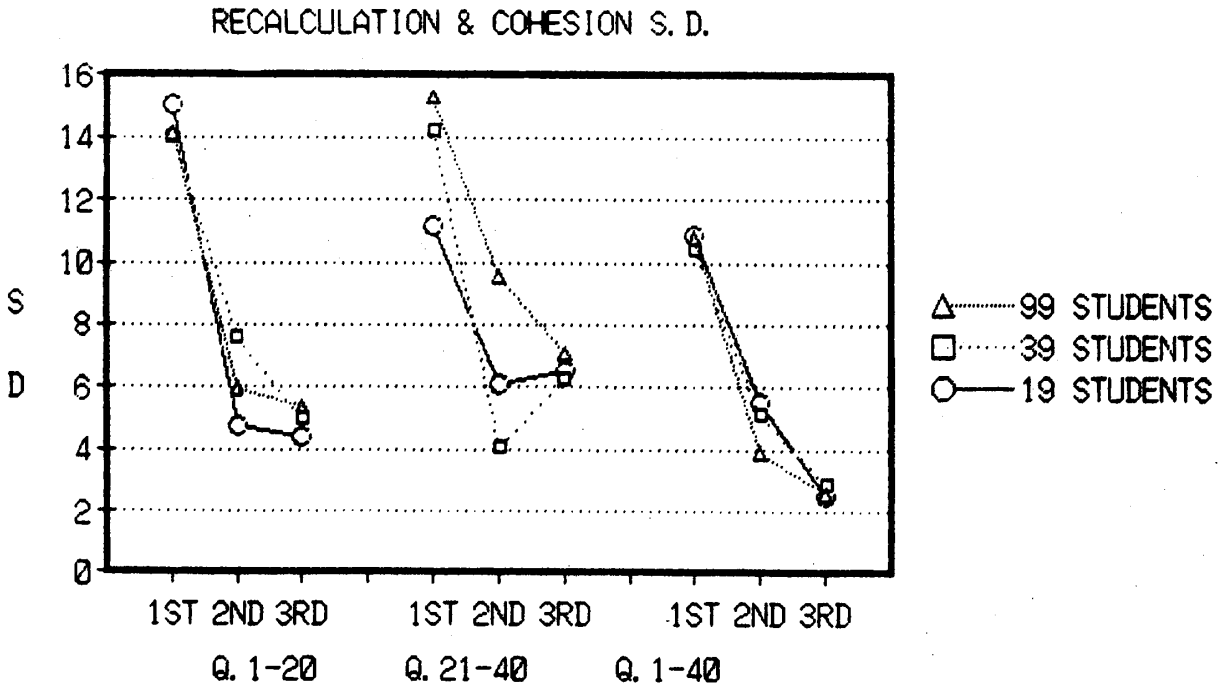
Table II

		COHESION QUOTIENTS											
		19	39	99	19	39	99	19	39	99	19	39	99
氏名	PTS.	ORIGINAL			RE-CALCULATED			FINAL			FINAL POINTS		
木	23	94	96	95	102	103	102	102	99	102	22	23	22
落	23	90	85	84	101	102	105	102	103	100	21	20	20
石	23	80	80	76	108	106	100	101	98	99	19	19	18
火	21	80	79	79	102	103	100	101	102	97	17	17	17
浦	19	92	88	89	104	101	103	99	102	101	18	17	18
合	19	87	87	88	101	99	103	98	100	103	17	17	17
大	19	85	86	85	100	102	103	101	103	97	16	16	17
河	18	86	84	81	105	106	100	99	99	99	16	16	15
紅	18	83	78	75	94	97	99	100	96	98	14	14	13
小	20	69	69	68	107	105	101	94	96	98	15	15	14
後	18	80	78	78	98	101	102	102	101	102	14	14	14
青	19	77	72	68	99	96	97	97	103	97	15	13	13
斧	17	82	78	78	95	102	107	106	104	100	13	13	14
伊	14	89	84	82	100	100	101	99	99	101	12	12	11
岡	17	72	67	67	89	101	103	101	103	103	11	11	11
愛	12	89	83	82	97	101	96	98	102	95	11	10	10
上	17	59	56	58	98	98	103	101	99	104	10	10	10
川	15	70	64	61	98	97	99	97	97	97	11	10	9
多	15	52	57	53	85	82	88	100	107	103	7	7	7
AVG.	18.3	81	78	77	99	101	101	99	100	99	14.7	14.4	14.2
S.D.		10.9	10.4	10.8	5.51	5.1	3.92	2.49	2.85	2.55			

It must be noted that the final scores are significantly affected by variation in test sample size: among the three, variations of zero to one point are predominant, although a variation of 3 points is not altogether uncommon, with an average maximum-minimum variation of approximately 1. Such variation is not so much a defect of the system as a natural result of fluctuation in question ordering.

To verify the cohesion approach from the question-sample size, the 40-question test was divided into two 20-question tests without regard to question response results. Improvement in cohesion quotients was confirmed, as was improvement in standard deviation (Figure II).

Figure II

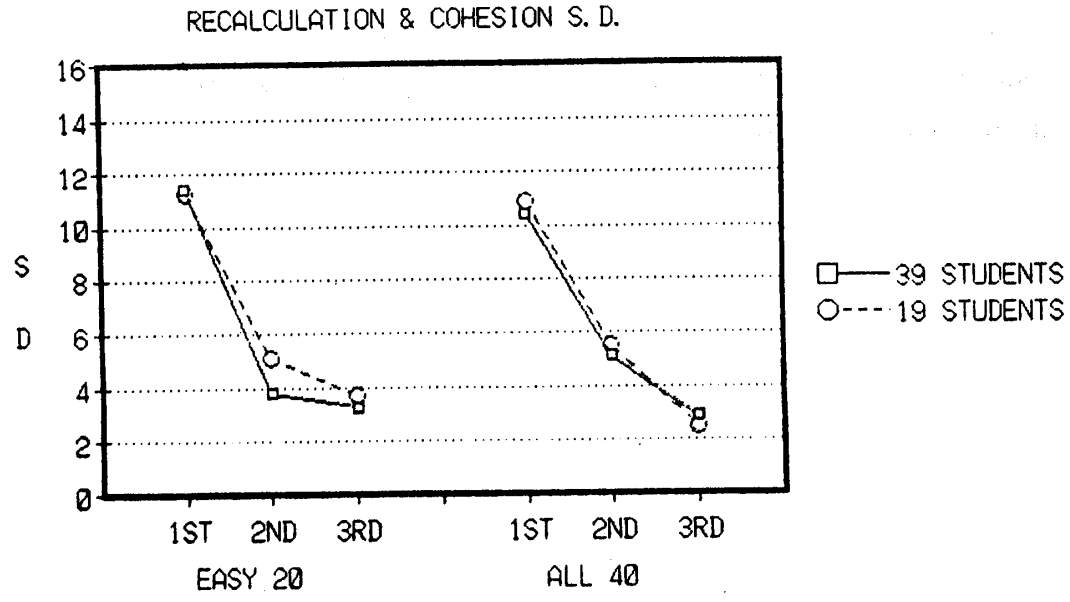


As a further verification of this cohesion approach, the 20 easiest questions were selected for test samples of 19 and 39 students. (Table III) (Note that doubling the test sample caused changes in the selection of the 20 questions involved.) Improvement in cohesion quotients was reconfirmed, while improvement in standard deviation was not dramatic as with the 40-question test.

Table III

	20 EASY QUESTIONS									
	19	19	39	39	19	39	19	39	19	39
	ORIGINAL				RE-CALC.		FINAL		FINAL PT.	
氏名	PT	CQ	PT	CQ	CQ	CQ	CQ	CQ	PT	PT
木	17	98	17	100	100	100	99	100	17	17
落	16	95	14	93	101	100	103	101	15	13
石	16	83	16	83	108	105	98	96	14	14
火	14	86	12	89	98	96	100	99	12	11
浦	15	98	12	99	98	100	99	99	15	12
合	16	90	14	93	103	98	103	101	14	13
大	15	88	16	89	105	104	96	97	14	15
河	12	96	13	91	105	101	96	101	12	12
江	14	86	15	79	96	102	104	99	11	12
小	10	83	11	77	105	106	102	110	8	8
後	14	84	14	82	96	108	108	98	11	12
青	15	80	12	77	101	105	100	102	12	9
斧	14	86	13	82	97	97	96	97	12	11
伊	11	96	11	90	91	101	103	99	10	10
岡	13	74	12	70	87	101	94	100	9	8
愛	10	95	9	92	95	105	106	104	9	8
上	11	64	10	62	96	104	98	104	7	6
川	9	84	9	75	94	102	96	99	8	7
多	10	54	12	57	98	92	99	105	5	6
AVG.	13.3	87	12.7	84	98	101	100	100	11.3	10.7
S.D.		11.2		11.4	5.1	3.79	3.68	3.25		

Figure III



These test samples clearly indicate the absence of any apparent inconsistency suggested by standard deviation.

Sample	Questions	Avg.	S. D.	Average Deviation
19	Q. 1 −20	7.53	4.42%	.33 points
39		7.42	4.97	.37
99		7.32	5.37	.39
19	Q. 21−40	7.21	6.50	.47
39		7.16	6.26	.45
99		6.84	7.06	.48
19	easy 20	11.3	3.68	.42
39		10.7	3.25	.34
19	Q. 1 −40	14.7	2.49	.37
39		14.4	2.85	.41
99		14.2	2.55	.36
				Average .40 points

The cohesion quotient approach has an average deviation of 0.40, but one which is composed of both numerical truncation-induced error and measurement uncertainty. Given a random sample of numbers between 0 and 0.99, the standard deviation of 0.29 would be found. Obviously the uncertainty of cohesion quotients (score reliability) is small indeed.

SUMMARY

The student/question measurement approach described is a potentially valuable tool in computer-scored testing, enabling an instructor to set aside the uncertainty of guessing, replacing it with an approximation confirmed to be superior to the traditional 1-point-per-question system. Just as it can be valuably utilized, it can also be dangerously misused: it is intended only for uniform instruction situations, such as single class testing—any time different instructors are involved, emphasis will be an unmeasured variable having potentially grave effects upon students' grades.